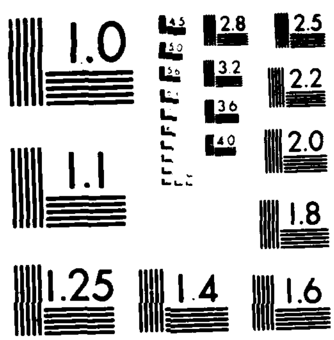
AD-A178 763 GOODNESS-OF-FIT BASED ON INTEGRATED SQUARED DISCREPANCIES BETWEEN SAMPLE A.. (U) RENSSELAUER POLYTECHNIC INST TROY NY SCHOOL OF MANAGEMENT 1/1

UNCLASSIFIED J L BRYANT ET AL. 1986 ARO-18072.19-MA F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

Rensselaer Polytechnic Institute

School of Management

Working Paper No. 37-86-P24

GOODNESS-OF-FIT BASED ON INTEGRATED SQUARED DISCREPANCIES BETWEEN SAMPLE AND POPULATION CHARACTERISTIC FUNCTIONS

by

J. L. Bryant

and

A. S. Paulson

School of Management
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
(518)266-6586

Note: This paper is not to be quoted without the permission of the author(s). For a list of Working Papers available from the School of Management, please contact Sheila Chao, School of Management.

GOODNESS-OF-FIT BASED ON INTEGRATED SQUARED DISCREPANCIES BETWEEN SAMPLE AND POPULATION CHARACTERISTIC FUNCTIONS

by

J.L. Bryant
University of Cincinnati

A.S. Paulson*
Rensselaer Polytechnic Institute

* Research sponsored in part by U.S. Army Research Office Contract No. DAAG29-81-K-0110

SUMMARY

We assess the feasibility of a goodness-of-fit test based on an intergral of the weighted squared modulus of the discrepancy between the sample and population characteristic functions. The resulting statistic is therefore analogous to the Cramer-von Mises statistic and is shown to reduce to it as a special case. A number of properties of the test have been derived, including the asymptotic null distribution of its statistic. It is shown that under mild regularity conditions the test is consistent. A number of approximations to the null distribution of the test statistic are considered, and are found to be successful in simplifying its application without due loss of accuracy.

Keywords: sample characteristic function, goodness-of-fit, weighted sum of chi-squared variates; consistency, asymptotic distribution; rate of convergence; cumulants, approximate distribution.

1. INTRODUCTION

The characteristic function corresponding to any distribution function $F(x)$ is defined by

$$\phi(u) = \int_{-\infty}^{\infty} e^{iux}\, dF(x) \tag{1.1}$$

where $i^2 = -1$, is uniquely associated with F (Lukacs, 1970, p. 28). For many purposes it is more convenient to work in this transform space rather than dealing directly with distribution functions. As a simple and well known example, the addition of independent random variables corresponds to a multiplication of characteristic functions. Similarly, the behavior of characteristic functions under location and scale shifts is particularly simple.

If $X_1, X_2, \ldots, X_n$ is a sample consisting of independent and identically distributed random variables with distribution function $F(x)$, then an empirical characteristic function $\hat{\phi}_n(u)$ may be defined as

$$\hat{\phi}_n(u) = \int_{-\infty}^{\infty} e^{iux}\, dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{iuX_j}. \tag{1.2}$$

and $\hat{\phi}_n(u)$ completely specifies the sample $X_1, X_2, \ldots, X_n$ up to permutations. In addition it follows almost by definition that, for each fixed u, $E(e^{iuX_j}) = \phi(u)$, so that $\hat{\phi}_n(u)$ is unbiased for $\phi(u)$. Thus by the strong law of large numbers (Rao, 1965, p. 97) $\hat{\phi}_n(u)$ is a strongly consistent estimator of $\phi(u)$ as well as being unbiased.

The potential applicability of sample characteristic functions in the inference setting of tests of goodness of fit has been investigated by Heathcote (1972), Paulson and Thornton (1975), Feurverger and Mureika (1977), Koutrouvelis (1980), Koutrouvelis and Kellermeier (1981), Murota and Takeuchi (1981), Csorgo and

Heathcote (1982), Bryant and Paulson (1982), Hall and Welsh (1983), Heathcote (1982), Epps and Pulley (1983), and Csorgo (1986). Further references and a review are given by Csorgo (1984).

This paper makes use of the goodness of fit statistic

$$n\Delta_n = n \int_{-\infty}^{\infty} |\hat{\phi}_n(u) - \phi_0(u)|^2 \, dw(u), \tag{1.3}$$

where $w(u)$ is a given weighting function. This staitstic is clearly analogous in form to the Cramer-von Mises statistic. There will be situations in which it is more natural to use (1.3) than the more standard empirical distribution function (EDF) statistics. For example, the existence of a viable characteristic function-oriented procedure would be of both theorectical and computational value in tests of hypotheses concerning the stable distributions, some compound distributions, or in cases where under the null hypothesis the population distribution is neither purely descrete nor absolutely continuous. Furthermore, most of our results are essentially independent of the dimension of the population under investigation and will be directly applicable to problems of multivariate goodness of fit. The multivariate case could be far more important than the univariate case because the uniform continuity of $\phi_0(u)$ and $\hat{\phi}_n(u)$ permits simplification in analysis of the test statistics which will not be available to multivariate EDF statistics.

2. A GOODNESS-OF-FIT TEST

The problem to be considered here is the construction of a procedure, based on the empirical characteristic function, whereby the goodness-of-fit hypothesis

$$H_0: \ F = F_0 \qquad \text{or} \qquad H_0: \ \phi(u) = \phi_0(u) \tag{2.1}$$

may be tested against the general alternative hypothesis

$$H_1: \; F \neq F_0 \qquad \text{or} \qquad H_1: \; \phi(u) \neq \phi_0(u). \tag{2.2}$$

Here F is the distribution function of the population in question and F_0 is completely specified; the corresponding characteristic functions are ϕ and ϕ_0 respectively. The formulation of *hypotheses* in *terms of* ϕ and ϕ_0 will be more natural for our purposes for we will be working for the most part in the transform space. The test procedure is based on the statistic (1.3) where $\hat{\phi}_n(u)$ is the empirical characteristic function computed from a random sample of size n, $X_1, X_2, \ldots, X_n$, drawn from the population. The weighting function $w(u)$ in (1.3) is a given distribution function, nondecreasing, continuous from the right and bounded, with $w(-\infty) = 0$, $w(\infty) = \beta > 0$. The value of the constant β is immaterial and will generally be taken to be unity. Since $|\hat{\phi}_n(u)|$ and $|\phi_0(u)|$ are bounded above by 1, the integral in equation (1.3) must converge. We will usually refer to w as a "weighting function" to avoid confusion with the distribution functions F and F_0.

The consistency of the test of null hypothesis based on $n\Delta_n$ is based on two results of Bryant and Paulson (1979).

<u>Lemma 2.1</u>

The quantity $\int_{-\infty}^{\infty} |\hat{\phi}_n(u) - \phi(u)|^2 \, dw(u)$, where $\hat{\phi}_n(u)$ is the empirical characteristic function based on a random sample of size n drawn from the distribution whose characteristic function is $\phi(u)$, has mean and variance given by

$$E \int_{-\infty}^{\infty} |\hat{\phi}_n(u) - \phi(u)|^2 \, dw(u) = \frac{1}{n} \int_{-\infty}^{\infty} \{1 - |\phi(u)|^2\} \, dw(u) \tag{2.4}$$

and

$$\mathrm{Var} \int_{-\infty}^{\infty} |\hat{\phi}_n(u) - \phi(u)|^2 \, dw(u) =$$

$$\frac{2(n-3)}{n^3} \left\{\int_{-\infty}^{\infty} |\phi(u)|^2 \, dw(u)\right\}^2 + \frac{n-1}{n^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{|\phi(u+t)|^2 + |\phi(u-t)|^2\} \, dw(u) dw(t)$$

$$- \frac{2(n-2)}{n^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{Re}\{\phi(u+t)\phi^*(u)\phi^*(t) + \phi(u-t)\phi^*(u)\phi(t)\} \, dw(u) dw(t). \tag{2.5}$$

Lemma 2.2

Let $n\Delta_n$ be defined as in equation (1.3) where $\hat{\phi}_n(u)$ is based on a random sample of size n from the population whose characteristic function is $\phi(u)$. Then Δ_n converges to the quantity $\int_{-\infty}^{\infty} |\phi(u) - \phi_0(u)|^2 \, dw(u)$ with probability one.

Since $w(u)$ is strictly increasing it places positive mass on every nondegenerate u-interval and this implies that the $\int_{-\infty}^{\infty} |\phi(u) - \phi_0(u)|^2 \, dw(u)$ is zero if, and only if, $F(x) = F_0(x)$. By Theorem 4.1 $n\Delta_n$ converges in law to a distribution function G. Thus the goodness-of-fit test based on $n\Delta_n$ will have a rejection rule of the form

Reject H_0 if and only if $n\Delta_n > c_a$"

where c_a is a critical value chosen so that asymptotically the size of the test is fixed at a.

If $F \neq F_0$, some x, than Δ_n will converge strongly to a constant other than zero as $n \to \infty$ so that $n\Delta_n$ will almost surely diverge up to ∞. This argument is summarized as

Theorem 2.1 The goodness-of-fit test of the hypothesis H_0: $F \equiv F_0$ based on the statistic Δ_n of (1.3), where the weighting function $w(u)$ is strictly increasing, is consistent; that is, for any alternate distribution $F \neq F_0$, the power of the test approaches one as

$n \to \infty$.

Tests of fit based on only a finite number of u-values cannot be consistent since the values taken on by a characteristic function at a finite number of u-values do not characterize a distribution function. The assumption in Theorem 2.1 that $w(u)$ be strictly increasing may be relaxed when the population of interest possesses an analytic characteristic function for then a characteristic function, and hence its distribution function is uniquely determined by its values over any non-degenerate u-interval (Lukacs, 1970, Chapter 7).

The statistic Δ_n is similar in structure and actually reduces as a special case to the Cramer-von Mises statistic

$$\omega_n^2 = \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 \, dF_0(x) \tag{2.6}$$

for testing H_0: $F = F_0$, where $F_0(x)$ is the completely specified distribution function and $F_n(x)$ is the sample distribution function corresponding to the random sample X_1, $X_2, \ldots, X_n$ putatively drawn from $F_0(x)$. We now sketch the proof of this assertion.

On the Hilbert space $L^2[0,1]$ of all square-integrable complex valued functions on $[0,1]$ the inner product of two functions h and g is given by $(f, g) = \int_0^1 f(t)g^*(t)dt$. The set of functions $\{e^{i2\pi kt}, k = 0,1,\ldots\}$ is a complete orthonormal system on $L^2[0,1]$.

By means of the probability integral transformation, (2.6) is reduced to the form

$$\omega_n^2 = \int_0^1 g_n^2(t) \, dt = ||g_n||^2 \tag{2.7}$$

where $g_n(t) = G_n(t) - t$ and $G_n(t)$ is the sample distribution function corresponding to the transformed sample $Y_1, Y_2, \ldots, Y_n$. By determining the Fourier coefficient of $g_n(t)$ with respect to $e^{i2\pi kt}$ and utilizing Parseval's equality in (2.7) it may be shown that

$$\omega_n^2 = \lim_{u \to 0} \frac{|\hat{c}_n(u) - c_0(u)|^2}{u} + \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \frac{|\hat{c}_n(-2\pi k) - c_0(-2\pi k)|^2}{4\pi^2 k^2} \tag{2.8}$$

$$= \int_{-\infty}^{\infty} |\hat{c}_n(u) - c_0(u)|^2 \, dw(u), \tag{2.9}$$

where $c_0(u)$ is the characteristic function of a random variable on [0,1] and $\hat{c}_n(u)$ is the empirical characteristic function associated with the Y_j and $w(u)$ is a step function whose definition is clear from equation (2.8). The Cramer-von Mises statistic is thus a special case of $n\Delta_n$.

3. ASYMPTOTIC DISTRIBUTION OF THE STATISTIC $n\Delta_n^{(p)}$

We first consider the distribution of

$$n\Delta_n^{(p)} = n \sum_{k=1}^{p} |\hat{\phi}_n(u_k) - \phi(u_k)|^2 \, w(u_k), \tag{3.1}$$

a discrete version of the $n\Delta_n$ of equation (1.3). The u_k in (3.1) are pre-selected abscissae and the $w_k = w(u_k)$ are given non-negative weights. There are at least two good reasons for considering $n\Delta_n^{(p)}$. First, it will turn out that the asymptotic distribution of $n\Delta_n$ may be obtained in essentially the same manner as that of $n\Delta_n^{(p)}$ of (3.1). Thus the simplified problem will serve to motivate later considerations. Secondly, there are situations of practical interest where $\phi(u)$ and $w(u)$ are such that the integral in equation (1.3) cannot be explicitly evaluated. If, for example, the desired weighting function were $dw(u) = \exp(-u^2)du$, then in equation (3.1) we might choose the u_k and w_k to be the abscissae and weights associated with the Hermitian quadrature of order p (Stroud and Secrist, 1966, p. 217).

We shall always take p to be even and $u_k = -u_{p-k+1}$, $w_k = w_{p-k+1}$, $k=1,2,\ldots,p$. Since $w_k = w_{p-k+1}$ and $(\mathrm{Re}\,\hat{\phi} - \mathrm{Re}\,\phi)(\mathrm{Im}\,\hat{\phi} - \mathrm{Im}\,\phi)$ is an odd function we may write (3.1) as

$$n\Delta_n^{(p)} = n \sum_{k=1}^{p} \{y_n(u) - y(u)\}^2 \, w_k. \tag{3.2}$$

where $y(u)$ is the transform, $y(u) = \mathrm{Re}\,\phi(u) + \mathrm{Im}\phi(u)$, and $\hat{y}_n(u)$ is its sample counterpart, $\hat{y}_n(u) = \mathrm{Re}\ \hat{\phi}_n(u) + \mathrm{Im}\ \hat{\phi}_n(u)$.

Some rearrangement gives

$$n\Delta_h^{(p)} = \left\{\left[n^{-\frac{1}{2}}\sum_{j=1}^{n} \underset{\sim}{s}_j\right]^T\left[n^{-\frac{1}{2}}\sum_{j=1}^{n} \underset{\sim}{s}_j\right]\right\} = \underset{\sim}{t}_n^T\, \underset{\sim}{t}_n \tag{3.3}$$

where

$$\underset{\sim}{t}_n = n^{-\frac{1}{2}}\sum_{j=1}^{n} \underset{\sim}{s}_j,\ \underset{\sim}{s}_j = (s_{1j}, s_{2j}, \ldots, s_{pj})^T,$$

and

$$s_{kj} = (\cos u_k X_j + \sin u_k X_j - \mathrm{Re}\ \phi(u_k) - \mathrm{Im}\ \phi(u_k))\, w_k^{\frac{1}{2}}.$$

The px1 vectors $\underset{\sim}{s}_j$ are independently and identically distributed with mean vector 0 and pxp covariance matrix $\underset{\sim}{M} = (m_{jk})$ given by

$$m_{jk} = K(u_j, u_k)(w_j w_k)^{\frac{1}{2}} \tag{3.4}$$

where, as is easily shown,

$$K(u,v) = \mathrm{Re}\ \phi(u-v) + \mathrm{Im}\ \phi(u+v) - [\mathrm{Re}\ \phi(u) + \mathrm{Im}\ \phi(u)][\mathrm{Re}\ \phi(v) + \mathrm{Im}\ \phi(v)]$$

$$= n \operatorname{cov}(\hat{y}_n(u), \hat{y}_n(v)).$$

Since $\underset{\sim}{M}$ is a covariance matrix it is symmetric and positive semi-definite of rank $r \le p$. We may therefore extract p real orthonormal eigenvectors $\underset{\sim}{c}_q$, $q = 1,2,\ldots, p$, which span p-dimensional Euclidean space and satisfy

$$\underset{\sim}{c}_q^T\, \underset{\sim}{M}\, \underset{\sim}{c}_k = \lambda_q \delta_{qk},\ 1 \le q,\ k \le p \tag{3.5}$$

where δ_{qk} is Kronecker's delta and $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the non-negative eigenvalues of $\underset{\sim}{M}$ which are arranged so that $\lambda_1, \lambda_2, \ldots, \lambda_r$ are positive while $\lambda_{p+1} = \ldots = \lambda_p = 0$. Referring the inner product of equation (3.3) to the orthonormal basis $\{\underset{\sim}{c}_1, \underset{\sim}{c}_2, \ldots, \underset{\sim}{c}_p\}$ gives

$$n\Delta_n^{(p)} = \sum_{q=1}^{r} (\underset{\sim}{c}_q^T \underset{\sim}{t}_n)^2 \tag{3.6}$$

since the random variables $\underset{\sim}{c}_q^T \underset{\sim}{t}_n$ are degenerate at the origin with probability one for $q > r$. From (3.6) and the definition of $\underset{\sim}{t}_n$,

$$n\Delta_n^{(p)} = \sum_{q=1}^{r} \lambda_q \left[\frac{1}{n^{\frac{1}{2}}} \sum_{j=1}^{n} z_{qj} \right]^2 \tag{3.7}$$

where

$$z_{qj} = \frac{\underset{\sim}{c}_q^T \underset{\sim}{g}_j}{\lambda_j^{\frac{1}{2}}}, \quad q = 1,2,\ldots, r; \; j = 1,2,\ldots, n.$$

The rx1 vectors $\underset{\sim}{z}_j = (z_{1j}, z_{2j}, \ldots, z_{rj})^T$, $j = 1,2,\ldots, n$, are independent and identically distributed with mean vector $\underset{\sim}{0}$ and, from (3.5), covariance matrix $\underset{\sim}{I}$. Since $n^{-\frac{1}{2}} \sum_{j=1}^{n} \underset{\sim}{z}_j$ converges in law to $N_r(0, \underset{\sim}{I})$ (Rao, 1965, p. 108) we find that $n\Delta_n^{(p)}$ is asymptotically distributed as

$$\sum_{q=1}^{r} \lambda_q \chi_q^2, \tag{3.8}$$

where χ_q^2 are independently and identically χ^2 distributed on one degree of freedom. The characteristic function of $n\Delta_n^{(p)}$ is

$$c^{(p)}(u) = \frac{\exp\left(\frac{i}{2} \sum_{q=1}^{r} \tan^{-1} (2\lambda_q u) \right)}{\prod_{q=1}^{r} (1+4\lambda_q^2 u^2)^{\frac{1}{4}}}, \quad \left| \sum_{q=1}^{r} \tan^{-1} 2\lambda_q u \right| < \frac{\pi}{2}, \tag{3.9}$$

$$= \{ \prod_{q=1}^{r} (1-2\lambda_q iu) \}^{\frac{1}{2}} \tag{3.10}$$

$$= |I - 2uiM|^{-\frac{1}{2}} \tag{3.11}$$

The proper branch of the square root in (3.10) is specified by (3.9). Equation (3.11) may be determined directly by observing that $n \sum_{j=1}^{p} (\hat{y}_n(u) -$

$y(u))^2$ is distributed asymptotically as a sum of correlated χ^2 variates (Lukacs and Laha, 1964, p. 40-43). We have thus proven

<u>Theorem 3.1</u>

The goodness-of-fit statistic $n\Delta_n^{(p)}$ defined by (3.1) has, when the null hypothesis H_0 is correctly specified, the asymptotic distribution whose characteristic function is

$$c^{(p)}(u) = \left(\prod_{q=1}^{r}(1-2\lambda_q iu)\right)^{-\frac{1}{2}} = |\underset{\sim}{I} - 2ui\underset{\sim}{M}|^{-\frac{1}{2}} \tag{3.12}$$

where $\underset{\sim}{M}$ is the covariance matrix whose elements are given by (3.4), $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the positive eigenvalues of $\underset{\sim}{M}$ and $r \leq p$ is the rank of $\underset{\sim}{M}$.

The determinant in (3.12) may be expanded (Pogorzelski, 1966, pp. 31-32) in terms of the covariance kernel and the weights w_j as

$$|\underset{\sim}{I} - \lambda \underset{\sim}{M}|^{-\frac{1}{2}} = 1-\lambda\sum_{j=1}^{p} K(u_j,u_j)w_j + \frac{\lambda^2}{2!}\sum_{j=1}^{p}\sum_{k=1}^{p} C(u_j,u_k)w_jw_k$$

$$- \ldots + \frac{(-\lambda)^p}{p!}\sum_{j_1=1}^{p}\cdots\sum_{j_1=1}^{p} C(u_{j_1},u_{j_2},\ldots,u_{j_p})\prod_{i=1}^{p} w_{j_i} \tag{3.13}$$

where we have set $\lambda = 2ui$ and

$$C(u_1,u_2,\ldots,u_s) = \begin{vmatrix} K(u_1,u_1) & \ldots & K(u_1,u_s) \\ \vdots & & \\ K(u_s,u_1) & \ldots & K(u_s,u_s) \end{vmatrix} . \tag{3.14}$$

Equations (3.12) and (3.13) are suggestive of the results that we may expect when we let p (as well as r) tend to infinity in such a way that (3.1) becomes the integral expression (1.3).

4. THE ASYMPTOTIC DISTRIBUTION OF $n\Delta_n$

This section contains our main result concerning the limiting distribution of the test statistic. Some preliminary lemmas are stated without proofs. The main reference for this section (and Section 6) is Dunford and Schwartz (1958).

Define the function

$$s_j(u) = \cos uX_j + \sin uX_j - \mathrm{Re}\ \phi(u) - \mathrm{Im}\ \phi(u) \tag{4.1}$$

for $j = 1,2,\ldots, n$. These processes are independently and identically distributed with mean function $E(s_j(u)) = 0$ and continuous covariance kernel $K(u,v)$ given at (3.4). Let $t_n(u)$ be defined by

$$t_n(u) = n^{-\frac{1}{2}} \sum_{j=1}^{n} s_j(u) = n^{\frac{1}{2}} (\hat{y}_n(u) - y(u)) \tag{4.2}$$

so that

$$n\Delta_n = \int_{-\infty}^{\infty} t_n^2(u)dw(u), \tag{4.3}$$

the squared norm of $t_n(u)$ in $L^2(w)$, the Hilbert space of functions square integrable with respect to $w(u)$. We shall employ Parseval's equality to expand $n\Delta_n$ of (4.3) in terms of a complete orthonormal system in $L^2(w)$ consisting at least partially of the eigenfunctions of the integral operator

$$K\ y(u) = \int_{-\infty}^{\infty} K(u,v)y(u)dw(v). \tag{4.4}$$

First we need some results concerning this operator.

<u>Lemma 4.1</u>

There exists a finite or countably infinite orthonormal system of eigenfunctions $f_q(u)$, $q = 1,2,\ldots,$ of the integral operator K of (4.4), where $f_q(u)$ is associated with the eigenvalue λ_q, which together with an at most countable set of functions $n_q(v)$

satisfying

$$\int_{-\infty}^{\infty} K(u,v)\, \eta_q(v)\, dw(v) = 0 \quad \text{a.e. } (w), \quad q = 1,2,\ldots,$$

form a complete orthonormal system in $L^2(w)$.

These assertions follow from the theory of linear operators of the Hilbert-Schmidt type in Hilbert space (Dunford and Schwartz, 1958, p. 1012, and Schmeidler, 1965, p. 57).

Lemma 4.2

The real and positive non-zero eigenvalues λ_1, $\lambda_2,\ldots,$ of the selfadjoint integral operator K of (4.4), arranged in decreasing order and repeated according to their multiplicity satisfy

$$\sum_{q=1}^{\infty} \lambda_q = \int_{-\infty}^{\infty} K(u,u)dw(u) = \int_{-\infty}^{\infty} (1 - |\Phi(u)|^2)dw(u) < \infty.$$

Lemma 4.3

The Fredholm determinant, defined by the power series

$$D(\lambda) = 1 + \sum_{j=1}^{\infty} \frac{(-\lambda)^j}{j!} \int_{R^j} C(u_1,u_2,\ldots, u_j)dw(u)\ldots dw(u_j),$$

where $C(u_1,u_2,\ldots,u_j)$ is defined in (3.14), is absolutely convergent for all complex λ and may be expressed as the infinite product

$$D(\lambda) = \prod_{q=1}^{\infty} (1-\lambda_q\lambda),$$

the latter converging absolutely and uniformly for all λ in any bounded region of the complex plane.

Because the kernel $K(u,v)$ is real, the real and imaginary parts of any eigenfunction $f_q(u)$ associated with the eigenvalue λ_q are also eigenfunctions associated with λ_q. The complete orthonormal basis $\{f_q(u),\ \eta_q(u),\ q = 1,2,\ldots\}$ of

Lemma 4.1 may therefore be assumed to be composed of real functions, and for simplicity we will take them as such.

Our primary result will be given; in the proof we will assume that the number of eigenvalues λ_q is infinite. If this is not the case (as in Theorem 3.1) the required alterations will be obvious.

Theorem 4.1

The goodness-of-fit statistic $n\Delta_n$ of equation (1.3) has, when the null hypothesis H_0 is correctly specified, the asymptotic distribution whose characteristic function $c(u)$ is

$$c(u) = \prod_{q=1}^{\infty} (1-2\lambda_q iu)^{-\frac{1}{2}} = D^{-\frac{1}{2}}(2ui)$$

where $\lambda_1, \lambda_2, \ldots,$ are the positive eigenvalues of the integral operator given by equation (4.4) with kernel $K(u,v)$ defined at (3.4). $D(\lambda)$ is the Fredholm determinant associated with this operator, defined in Lemma 4.3.

Proof

Using Parseval's equality, we expand the squared norm in equation (4.3) with respect to the complete orthonormal system $\{f_q(u), \eta_q(u), q = 1,2,\ldots\}$. This gives

$$n\Delta_n = \sum_{q=1}^{\infty} (t_n, f_q)^2 + \sum_{q=1}^{\infty} (t_n, \eta_q)^2 \ . \tag{4.5}$$

Now by Lemma 3.1 we have $E(t_n(u)) = 0$ and $\mathrm{Cov}(t_n(u), t_n(v)) = K(u,v)$. Therefore, use of Fubini's theorem gives, for $q = 1,2,\ldots,$

$$E(t_n, \eta_q) = E(t_n, f_q) = 0,$$

$$E\{(t_n, \eta_q)^2\} = 0, \tag{4.6}$$

$$E\{(t_n, f_q)^2\} = \lambda_q \int_{-\infty}^{\infty} f_q^2(u) = \lambda_q. \tag{4.7}$$

From equations (4.6) and (4.7) we have

$$\sum_{q=1}^{\infty} E(t_n,\ f_q)^2 + \sum_{q=1}^{\infty} E(t_n,\ \eta_q)^2 = \sum_{q=1}^{\infty} \lambda_q$$

which we have shown to be finite. Accordingly (Rao, 1965, p. 91), the sum in equation (4.5) is convergent with probability one and $E(n\Delta_n) = \sum_{q=1}^{\infty} \lambda_q$.

Each of the terms $(t_n,\ \eta_q)^2$ in equation (4.5) is, with probability one, degenerate at the origin. Since there are at most a countable number of such terms, we have that with probability one $n\Delta_n$ may be expressed as

$$n\Delta_n = \sum_{q=1}^{\infty} (t_n,\ f_q)^2. \tag{4.8}$$

As in the proof of Theorem 3.1, let

$$z_{q,j} = \frac{(s_j,\ f_q)}{\lambda_q^{\frac{1}{2}}},\ q = 1,2,\ldots,\quad j = 1,2,\ldots,n,$$

so that

$$n\Delta_n = \sum_{q=1}^{\infty} \lambda_q \left[\frac{1}{n^{\frac{1}{2}}} \sum_{j=1}^{n} z_{q,j}\right]^2 . \tag{4.9}$$

Thus, $E(z_{qj}) = 0$ and

$$\begin{aligned}\mathrm{Cov}(z_{qj},\ z_{q'j}) &= \frac{1}{(\lambda_q\lambda_{q'})^{\frac{1}{2}}} \int_{-\infty}^{\infty} f_q(u) \left[\int_{-\infty}^{\infty} K(u,v) f_{q'}(v)dw(v)\right] dw(u) \\ &= \frac{\lambda_{q'}}{(\lambda_q\lambda_{q'})^{\frac{1}{2}}} \int_{-\infty}^{\infty} f_q(u) f_{q'}(u)dw(u) = \delta_{qq'} .\end{aligned}$$

If we now define the truncated sums

$$n\Delta_n^{(r)} = \sum_{q=1}^{r} (t_n,\ f_q)^2 = \sum_{q=1}^{r} \lambda_q (n^{-\frac{1}{2}} \sum_{j=1}^{n} z_{qj})^2 \tag{4.10}$$

$r = 1,2,\ldots,$ it follows by Theorem 3.1 that $n\Delta_n^{(r)}$ is asymptotically distributed as $Q^{(r)} = \sum_{q=1}^{r} \lambda_q \chi_q^2$, a weighted sum of independent χ^2 variates one degree of freedom, with characteristic function $c^{(r)}(u)$ given by (3.10) with p=r.

Let F_n be a sequence of distribution function converging to a distribution function F on the continuity set of F. This mode of convergence is denoted by $F_n \Rightarrow F$. Let G_n be the distribution function of $n\Delta_n$, $G_n^{(r)}$ the distribution function of $n\Delta_n^{(r)}$ and $G^{(r)}$ the distribution function of $Q^{(r)}$ whose characteristic function is $c^{(r)}(u)$. We have shown that $G_n^{(r)} \Rightarrow G^{(r)}$ for fixed r. An application of the continuity theorem (Lukacs, 1970, pp. 48-50) and Lemma 4.3 implies that $G^{(r)} \Rightarrow G$.

The proof is completed by constructing a sequence $\{r_n, n = 1,2,\ldots\}$ which diverges monotonically up to $+\infty$ and showing that $G_n^{r_n} \Rightarrow G$ which will imply that $G_n \Rightarrow G$ (Rao, 1965, pp. 100-102). Since the distribution functions $G^{(r)}(x)$ are continuous, we have by Polya's theorem (Rao, 1965, p. 100)

$$\lim_{n\to\infty} \sup_x |G_n^{(r)}(x) - G^{(r)}(x)| = 0.$$

Hence we can define integers q_p by q_p = least integer $> q_{p-1}$ such that $|G_n(x)^{(p)}\ G^{(p)}(x)| < 2^{-p}$ for all $n \geq q_p$, with $q = 1,2,\ldots,$ $q_0 = 0$. We than take

$$r_n = p \text{ for } n = q_p + 1,\ q_p + 2, \ldots,\ q_{p+1},$$

$n = 1,2,\ldots,$ $p = 1,2,\ldots$

Then for any x and p

$$|G_n^{(r_n)}(x) - G^{(p)}(x)| < \frac{1}{2^p}\ ,\ n = q_p+1,\ q_p+2, \ldots,\ q_{(p+1)} \qquad (4.11)$$

Let $\epsilon>0$ be given, and let x be any point of G. Since $G^{(r)} \Rightarrow G$, there exists an r^* such that

$$|G^{(r)}(x) - G(x)| < \frac{\epsilon}{2} \quad \text{for all } r \geq r^*. \tag{4.12}$$

Without loss of generality, take r^* large enough so that $\frac{1}{2^{r^*}} < \frac{\epsilon}{2}$.

Let $n^* = q_{r^*}+1$. Then for any $n \geq n^*$, there exits an $\bar{r} \geq r^*$ such that $q_{\bar{r}} + 1 \leq n \leq q_{(\bar{r}+1)}$. For such n we have, by equation ( )

$$|G_n^{(r_n)}(x)-G^{(\bar{r})}(x)| < \frac{1}{2^{\bar{r}}} \leq \frac{1}{2^{r^*}} < \frac{\epsilon}{}$$

while from equation (4.12),

$$|G^{(\bar{r})}(x) - G(x)| < \frac{\epsilon}{2} .$$

Thus

$$|G_n^{(r_n)}(x) - G(x)| \leq |G_n^{(r_n)}(x) - G^{(\bar{r})}(x)| + |G^{(\bar{r})}(x) - G(x)| < \epsilon$$

for all $n \geq n^*$. It follows that $G_n^{(r_n)} \Rightarrow G$, which in turn implies that $G_n \Rightarrow G$, and so the theorem is proven.

The correspondence between Theorems 3.1 and 4.1 is clear. Although it was assumed in Theorems 3.1 and 4.1 that the underlying population was univariate, the proof would still go through with only notational changes if the populations were multivariate. Note that the function w(u) need not be absolutely continuous and the integral in the definition of the $n\Delta_n$ need not be restricted to a finite support. For example, the symmetry statistic nT_n of Feuerverger and Mureika (1977) is of the same general form as $n\Delta_n$ and Theorem 3.1 may be used to derive the asymptotic distribution of this statistic under the null hypothesis of population symmetry.

5. SPECIAL CASES AND EXAMPLES

The results of Theorem 4.1 will be illustrated in this section by applying them in several specific examples.

(a) Suppose the hypothesized distribution is discrete and that under H_0 the population random variable may take on the values $x_1, x_2, \ldots, x_m$ with positive probabilities $p_1, p_2, \ldots, p_m$, $\sum_{q=1}^{m} p_q = 1$. Then $\mathrm{Re}\,\phi(u) = \sum_{q=1}^{m} p_q \cos(ux_q)$ and $\mathrm{Im}\phi(u) = \sum_{q=1}^{m} p_q \sin(ux_q)$. A little algebra will reduce the kernel $K(u,v)$ at equation (3.4) to

$$K(u,v) = \sum_{q=1}^{m} g_q(u) g_q(v) p_q \ ,$$

where

$$g_q(u) = \cos(ux_q) + \sin(ux_q) - \mathrm{Re}\phi(u) - \mathrm{Im}\phi(u).$$

Thus the kernel is, in this case, degenerate and it may be shown by an argument similar to that given by Smithies (1958, pp. 36-38) that the nonzero eigenvalues of $K(u,v)$ are precisely those of the matrix product $\underset{\sim}{A}\underset{\sim}{P}$, where the $(j,k)^{th}$ element of $\underset{\sim}{A}$ is

$$a_{jk} = \int_{-\infty}^{\infty} g_j(u) g_k(u) dw(u) \quad 1 \le j,\ k \le m,$$

and $\underset{\sim}{P} = \mathrm{diag}\ \{p_1, p_2, \ldots, p_m\}$.

It is seen that the situation here is similar to the one considered in Theorem 3.1. The asymptotic distribution of $n\Delta_n$ is that of a finite sum of independent weighted chi-squared variates, the weights being the nonzero eigenvalues of $\underset{\sim}{A}\underset{\sim}{P}$.

The case described above applies whenever data have been grouped; it also indicates a possible approximation method to be used when the eigenvalues of the kernel $K(u,v)$ are not easily obtained.

(b) Suppose that under the null hypothesis the underlying population has a Cauchy distribution with location and scale parameters δ and σ respectively, so that the population characteristic function is $\Phi(u) = \exp(i\delta u - \sigma|u|)$.

Suppose the weight function

$$w(u) = \alpha \int_{-\infty}^{u} e^{-\alpha|t|}dt$$

is chosen, where α is a given positive constant. Then $n\Delta_n$ may be explicitly integrated and is given by

$$n\Delta_n = \frac{1}{n} \sum_{j,k=1}^{n} \frac{2}{1+\left[\frac{X_j - X_k}{\alpha}\right]^2} - \sum_{j=1}^{n} \frac{4(1+\sigma/\alpha)}{(1+\sigma/\alpha)^2 + \left[\frac{X_j - \delta}{\alpha}\right]^2} + \frac{2n}{1 + 2\sigma/\alpha} \qquad (5.1)$$

where $X_1, X_2, \ldots, X_n$ is the random sample drawn from the population. The variates $\frac{X_j - \delta}{\alpha}$ have, if the null hypothesis is true, the Cauchy distribution with location parameter zero and scale parameter $\gamma = \frac{\sigma}{\alpha}$. Therefore, the null distribution of (5.1) depends only on γ, which we will here take to be unity for simplicity. We may thus assume that $\Phi(u) = e^{-|u|}$ and $w(u) = \int_{-\infty}^{u} e^{-|t|}dt$. Then the covariance kernel $K(u,v)$ (see 3.4) is equal to

$$K(u,v) = e^{-|u-v|} - e^{-|u|-|v|} . \qquad (5.2)$$

It may be shown that the asymptotic characteristic function of $n\Delta_n$ is given by

$$c(u) = D^{-\frac{1}{2}}(2ui) = \prod_{s=1}^{\infty} \left[1 - \frac{16iu}{j_{2s}^2}\right]^{-1} \qquad (5.3)$$

where j_{2s}, $s = 1,2,\ldots$, are the positive zeros of $J_2(u)$, the Bessel function of the first kind and order 2. Equation (5.3) characterizes the asymptotic distribution of $n\Delta_n$ as that of an infinite sum of exponential deviates having means $16/j_{2s}^2$. The zeros j_{2s} have been tabulated for $1 \leq s \leq 20$ in Abramowitz and Stegun (1970,

p. 409).

(c) The ability to analytically obtain the eigenvalues associated with the asymptotic distribution of $n\Delta_n$, as in the previous example, is unfortunately rather unusual. A more typical sitaution is illustrated by assuming that the hypothesized distribution is normal with mean δ and variance σ^2. The postulated population characteristic function is then $\phi(u) = \exp(i\delta u - \frac{1}{2}\sigma^2u^2)$. If $dw(u) = \alpha \exp(-\alpha^2u^2)\, du$, $\alpha > 0$, is chosen, then equation (1.3) may be integrated to yield

$$n\Delta_n = \frac{\sqrt{\pi}}{n} \sum_{j,k=1}^{n} \exp(-(X_j-X_k)^2/4\alpha^2) - \frac{2\sqrt{2\pi}\,\alpha}{(\sigma^2+2\alpha^2)^{\frac{1}{2}}} \sum_{j=1}^{n} \exp(-(X_j-\delta)^2/(2\sigma^2+4\alpha^2))$$

$$+ \frac{n\sqrt{\pi}\,\alpha}{(\sigma^2+\alpha^2)^{\frac{1}{2}}} . \tag{5.4}$$

It is possible to obtain the asymptotic characteristic function of the statistic $n\Delta_n$ of equation (5.4) under a correctly specified null hypothesis. As an example, when α is chosen equal to σ the asymptotic characteristic function $c(u)$ of $n\Delta_n$ may be shown to be equal to

$$\prod_{q=0}^{\infty}(1-2bg^q iu)^{-\frac{1}{2}} \{1+2diu \sum_{m=0}^{\infty} \begin{bmatrix} -\frac{1}{2} \\ m \end{bmatrix} (-f)^m/(1-2bg^{2m}iu)\}^{-\frac{1}{2}} \tag{5.5}$$

where b, d, f and g are certain constants whose values are not important to the discussion. Unfortunately the characteristic function $c(u)$ is not readily inverted numerically, and so this result seems to be of little practical value. What are really needed are the values of the weights associated with the representation of the asymptotic distribution of $n\Delta_n$ as a weighted sum of independent chi-squared random variables, and these are not made evident by formula 5.5.

Fortunately, it turns out that the asymptotic distribution of $n\Delta_n$ in the case where the underlying population is normal may be accurately approximated by the asymptotic distribution of a quadrature-type sum of the form considered in Theorem 3.1. Such an approximate procedure is easier and more accurate than the numerical inversion of the expression in equation (5.5). As in example (b) the null distribution of (5.4) depends only on the ratio $\gamma = \sigma/\alpha$, so we may take without loss of generality $\phi(u) = \exp(-\frac{1}{2}\gamma^2 u^2)$ and $dw(u) = \exp(-u^2)du$ so that

$$n\Delta_n = n \int_{-\infty}^{\infty} |\hat{\phi}_n(u) - \phi(u)|^2 e^{-u^2} du. \tag{5.6}$$

The form of equation (5.6) suggests that the asymptotic distribution of $n\Delta_n$ be approximated by that of (3.1) where u_k and w_k, $k = 1,2,\ldots, p$, are respectively the zeros of the Hermite polynomial of (even) order p and the associated Hermitian quadrature weights. These have been extensively tabulated by Stroud and Secrist (1966, p. 217). The asymptotic characteristic function of $n\Delta_n^{(p)}$ is given in Theorem 3.1, where $\lambda_q^{(p)}$, $q = 1,2,\ldots, p$, are the eigenvalues of the $p\times p$ matrix $M^{(p)}$ whose $(q, q')^{th}$ element is, from (3.4)

$$m_{qq'}^{(p)} = (e^{-\gamma^2(u_q-u_{q'})^2/2} - e^{-\gamma^2 u_q^2/2-\gamma^2 u_{q'}^2/2})(w_q w_{q'})^{\frac{1}{2}} \tag{5.7}$$

To aid in the comparison of the distribution of $n\Delta_n$ and $n\Delta_n^{(p)}$, we tabulated the values of the first four cumulants of the asymptotic distributions of $n\Delta_n^{(p)}$ and of $n\Delta_n$ for various values of p and for γ ranging from $\frac{1}{2}$ to 3. The j^{th} cumulant of $n\Delta_n^{(p)}$ is given by

$$\kappa_j^{(p)} = (j-1)!2^{j-1}\sum_{q=1}^{p}(\lambda_q^{(p)})^j, \quad j = 1,2,\ldots, \tag{5.8}$$

which follows from the characterization of $n\Delta_n^{(p)}$ as a weighted sum of independent chi-squared random variables each with one degree of freedom, along with the additivity of cumulants. The cumulants κ_j, $j=1,2,\ldots$, of $n\Delta_n$ are calculated by formulas which will be provided in Section 7. In particular κ_1 and κ_2 can be obtained by the results of Lemma 2.1.

Our computations indicate that the first four cumulants of $n\Delta_n^{(p)}$ reach their asymptotic values for $\gamma = \frac{\sigma}{\alpha} = \frac{1}{2}$ when p=12, for γ=1 when p=20, for γ=3/2 when p=36, and for γ=3 when p=64. The cumulants of $n\Delta_n^{(p)}$ are rapidly convergent to those of $n\Delta_n$ if the ratio $\gamma = \sigma/\alpha$ is not large. When γ is large the kernel $K(u,v)$ becomes poorly behaved and is therefore difficult to approximate by $\underset{\sim}{M}$. For γ relatively small and p sufficiently large the asymptotic distribution of $n\Delta_n^{(p)}$ provides a good approximation to the distribution of $n\Delta_n$.

The use of approximating sums of the form given in equation (3.1) to determine the asymptotic distribution of $n\Delta_n$ is not always as effective as the preceding results might lead one to suspect. Referring back to example (b) where the hypothesized population was Cauchy with scale parameter one and the weighting function $dw(u) = \exp(-|u|)du$ was employed in the definition of $n\Delta_n$, it is reasonable to approximate $n\Delta_n$ by use of a quadrature sum of the Laguerre type (Stroud and Secrist, 1966, p. 253). The asymptotic distributions of such sums, however, converge to that of $n\Delta_n$ rather more slowly than was seen in the case where the underlying population was normal. This is to be expected, since the covariance kernel $K(u,v)$ corresponding to the Cauchy distribution, given in equation (5.2), is not everywhere differentiable. Thus by approximating the distribution of $n\Delta_n$ by that of a statistic of the nature of $n\Delta_n^{(p)}$ of equation (3.1), we are essentially replacing the problem of obtaining the positive eigenvalues associated with a homogeneous integral equation of the form

$$\lambda y(u) = \int_{-\infty}^{\infty} K(u,v)y(v)dw(v) \tag{5.9}$$

with the problem of finding those corresponding to the linear system

$$\lambda y_q = \sum_{q'=1}^{p} K(u_q,u_{q'})w_{q'}y_{q'} \quad q = 1,2,\ldots, p. \tag{5.10}$$

It cannot be expected that the linear system accurately approximates the integral equation unless the kernel $K(u,v)$ is smooth.

6. INVERSION OF THE CHARACTERISTIC FUNCTION

The distribution of the goodness-of-fit statistic $n\Delta_n$ is given in terms of its characteristic function $c(u)$ by Theorem 4.1, but in order to apply this statistic to a given testing situation, an inversion of the characteristic function is required to produce the necessary critical values. The problem of inverting $c(u)$ is similar to that considered by Durbin and Knott (1972) in their analysis of the residual sums B_p of the Cremer-von Mises statistic and the method of solution outlined here is essentially identical to their procedure, which is based on the work of Gil-Pelaez (1951) and Imhoff (1961).

The asymptotic characteristic function of $n\Delta_n$ is, from Theorem 4.1

$$c(u) = \prod_{q=1}^{\infty} (1-2\lambda_q iu)^{-\frac{1}{2}} = \frac{\exp\left\{\frac{1}{2} \sum_{q=1}^{\infty} \tan^{-1}(2\lambda_q u)\right\}}{\prod_{q=1}^{\infty} (1+4\lambda_q^2 u^2)^{\frac{1}{4}}} . \tag{6.1}$$

Denoting by $G(x)$ the asymptotic distribution function of $n\Delta_n$ and making the substitution $2t = u$ in the inversion formula of Gil-Pelaez (1951), leads to the approximation (Imhoff, 1961)

$$G(x) \cong \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\sin\left\{\frac{1}{2} \sum_{q=1}^{p} \tan^{-1}(\lambda_q u) - \frac{1}{2}ux\right\}}{u \prod_{q=1}^{p} (1 + \lambda_q^2 u^2)^{\frac{1}{4}}} \, du \tag{6.2}$$

where λ_p is adjusted by setting it equal to $E(n\Delta_n) - \sum_{q=1}^{p-1} \lambda_q$ so that a match of the expectations of the two sums is effected. The integral in (6.2) is evaluated by Gaussian quadrature with repeated interval halving until the error, based on successive iterations, is acceptably small. Expression (6.2) can be used to compute the exact values of $G(x) = \Pr\{n\Delta_n \le x\}$.

7. APPROXIMATIONS TO THE DISTRIBUTION OF $n\Delta_n$

The results of Sections 2, 3, 4, and 5 make possible the application of the statistic $n\Delta_n$ to tests of goodness-of-fit; its distribution under a correctly specified null hypothesis has been ascertained in terms of its characteristic function, and a procedure by which this characteristic function may be inverted has been presented, so that the critical values on which the hypothesis test is based may be calculated. Even so, from a practical point of view these results are unsatisfactory; it has been seen that, in order to obtain the characteristic function of $n\Delta_n$, the eigenvalues of a certain integral operator must be determined, and once this has been done a numerical integration is required to perform the necessary inversion. (This is not the case for Theorem 3.1.) Of course, if the proposed testing procedure were to be used only in a limited number of more or less typical situations this would cause no difficulty, for then tables of the critical values of $n\Delta_n$ relevant to those situations could be prepared. A goodness-of-fit test based on the empirical characteristic function should, however, be applicable to a wide range of problems precisely because it is in unusual circumstances, where the standard tests are not applicable or else can only be employed with difficulty, that such a testing procedure would be of greatest value.

It is clear that reasonably simple and yet accurate approximations to the distribution of $n\Delta_n$ are required. It will now be shown that certain approximations, formulated in terms of the cumulants of $n\Delta_n$, meet these criteria. Simulation results presented in this section indicate that the convergence of the distribution of $n\Delta_n$ is quite rapid, implying that our results are of practical use for moderate sample sizes.

The iterated kernels $K_j(u,v)$ associated with the kernel $K(u,v)$ are recursively defined by $K_1(u,v) = K(u,v)$ and

$$K_j(u,v) = \int_{-\infty}^{\infty} K_{j-1}(u,t)K(t,v)dw(t), \quad j \geq 2.$$

It is shown in Dunford and Schwartz (1963, pp. 1085-7) that

$$\beta_j = \int_{-\infty}^{\infty} K_j(u,u)dw(u) = \sum_{q=1}^{\infty} \lambda_q^j, \text{ for } j \geq 1,$$

and

$$D(\lambda) = \exp\left\{- \sum_{j=1}^{\infty} \beta_j \lambda^j / j\right\}$$

for all $|\lambda|$ sufficiently small. Since from Theorem 4.1 the asymptotic characteristic function $c(u)$ of $n\Delta_n$ is given by $D^{-\frac{1}{2}}(2ui)$,

$$c(u) = \exp\left\{\tfrac{1}{2} \sum_{j=1}^{\infty} \beta_j (2ui)^j / j\right\}.$$

Finally, this gives the cumulant generating function, valid for all $|u|$ sufficiently small (Cramer, 1946, pp. 185-6),

$$\log c(u) = \tfrac{1}{2} \sum_{j=1}^{\infty} \beta_j (2ui)^j / j = \sum_{j=1}^{\infty} \kappa_j (iu)^j / j!.$$

Equating coefficients of u^j then yields

$$\kappa_j = (j-1)!\ 2^{j-1} \beta_j \ . \tag{7.1}$$

If the weighting function $w(u)$ in equation (1.3) is chosen so that $n\Delta_n$ can be explicitly integrated, as in examples (b) and (c) of Section 5, then the evaluation of the first several integrals in β_j may often be performed by straightforward, if somewhat tedious, calculations. We will subsequently need expressions for the asymptotic cumulants so derived for (b) and (c) of Section 5. We record the results for the Gaussian case only.

(a) If the underlying population is normal with mean δ and variance σ^2, and $dw(u) = \alpha \exp(-\alpha^2 u^2)du$, $\alpha > 0$, then, putting $\gamma = \frac{\sigma}{\alpha}$,

$$\kappa_1 = \sqrt{\pi}\,\{1-(\gamma^2+1)^{-\frac{1}{2}}\} \tag{7.2a}$$

$$\kappa_2 = 4\pi\{(8\gamma^2+4)^{-\frac{1}{2}} - 2(3\gamma^4+8\gamma^2+4)^{-\frac{1}{2}} + 2(\gamma^2+1)^{-1}\} \tag{7.2b}$$

$$\kappa_3 = 16\pi^{3/2}\{(9\gamma^4+12\gamma^2+4)^{-\frac{1}{2}} - 3(2\gamma^6+10\gamma^4+12\gamma^2+4)^{-\frac{1}{2}} \tag{7.2c}$$

$$+ 3(3\gamma^6+11\gamma^4+12\gamma^2+4)^{-\frac{1}{2}} - (4\gamma^6+12\gamma^4+4)^{-\frac{1}{2}}\}$$

$$\kappa_4 = 192\pi^2\{[4(\gamma^2+1)(2\gamma^2+1)^{\frac{1}{2}}]^{-1} - 4[(3\gamma^2+2)^2(\gamma^2+2)^2 - 4\gamma^4(\gamma^2+1)^2]^{-\frac{1}{2}}$$

$$+ 2[(\gamma^2+1)(4(\gamma^2+1)^2-2\gamma^4)^{\frac{1}{2}}]^{-1} + 2[(3\gamma^2+2)(\gamma^2+2)]^{-1} \tag{7.2d}$$

$$- 2[(\gamma^2+1)((3\gamma^2+2)(\gamma^2+2))^{\frac{1}{2}}]^{-1} + [2(\gamma^2+1)]^{-2}\}$$

It seems natural to approximate the null distribution of $n\Delta_n$ with that of a weighted chi-squared variate, or perhaps the sum of several such random variables, the weights and degrees of freedom associated with the approximating variates being chosen in such a way that the first several cumulants of $n\Delta_n$ are matched. Three such procedures have been investigated:

(a) <u>Patnaik's (1949) two-cumulant χ^2 approximation</u>

The null distribution of $n\Delta_n$ is approximated by that of the variate $Q = a\chi^2_\nu$; a and ν are determined by matching the first two cumulants of $n\Delta_n$ with those of Q. This requires $a = \kappa_2/2\kappa_1$, $\nu = 2\kappa_1^2/\kappa_2$. Then, given any $x > 0$,

$$P(n\Delta_n \le x) \cong P(\chi^2 \le \frac{2x\kappa_1}{\kappa_2}).$$

(b) Pearson's (1959) three-cumulant χ^2 approximation

The approximating variate has the form

$$W = (\chi^2 - \nu)\frac{\kappa_3}{4\kappa_2} + \kappa_1,$$

where here χ^2 is a chi-squared variate with $\nu = 8\kappa_2^3/\kappa_3^2$ degrees of freedom so that $n\Delta_n$ and W have the same first three cumulants. Thus

$$P(n\Delta_n \le x) \cong P(\chi^2 \le y), \quad y = (x-\kappa_1)\frac{4\kappa_2}{\kappa_3} + \nu .$$

(c) Sum of two independent weighted chi-squared variates

Let the approximating random variable be

$$Z = a_1\chi_1^2 + a_2\chi_2^2 \tag{7.3}$$

where χ_1^2 and χ_2^2 are independent and have chi-squared distributions with ν_1 and ν_2 degrees of freedom respectively. Equating the first four cumulants of Z and $n\Delta_n$ requires that

$$\begin{aligned} a_1\nu_1 + a_2\nu_2 &= \kappa_1, & 2a_1^2\nu_1 + 2a_2^2\nu_2 &= \kappa_2, \\ 8a_1^3\nu_1 + 8a_2^3\nu_2 &= \kappa_3, & 48a_1^4\nu_1 + 48a_2^4\nu_2 &= \kappa_4 . \end{aligned} \tag{7.4}$$

Some algebra will show that the values of a_1, a_2, ν_1 and ν_2 satisfying (7.4) may be obtained by taking a_1 and a_2 to be the solutions of the quadratic equation

$$(\beta_2^2 - \beta_1\beta_3)a^2 + (\beta_1\beta_4 - \beta_2\beta_3)a + (\beta_3^2 - \beta_2\beta_4) = 0 \tag{7.5}$$

with

$$\nu_1 = \frac{\beta_1 a_2 - \beta_2}{a_1(a_2 - a_1)}, \quad \nu_2 = \frac{\beta_2 - \beta_1 a_1}{a_2(a_2 - a_1)}; \tag{7.6}$$

and β_i determined from (7.1).

In every situation to which this approximation has been applied, equations (7.5) and

(7.6) have yielded real and positive values for a_1, a_2, ν_1 and ν_2.

Direct inversion of the characteristic function of Z, in (7.3), gives, for $\nu_1 + \nu_2 > 2$ (Gradshteyn and Ryzhik, 1965, p. 320, 3.394-7), an expression for the density of Z which may be integrated termwise and rearranged to yield the distribution function $G_Z(x)$ of Z as

$$G_Z(x) = \frac{(a_1/a_2)^{\frac{1}{2}\nu_2}}{\Gamma(\frac{1}{2}\nu_2)} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}(\nu_2+2k))}{k!} \left[1 - \frac{a_1}{a_2}\right]^k V_{\nu_1+\nu_2+2k} \left[\frac{x}{a_1}\right] \tag{7.7}$$

where $V_\nu(\cdot)$ is the distribution function of a chi-squared variate with ν degrees of freedom and a_1 is the smaller of the weights a_1 and a_2. The above development allows one to approximate the distribution of $n\Delta_n$ by $P(n\Delta_n \le x) \cong G_Z(x)$.

If the underlying population is specified to be normal with mean 8 and variance σ^2 under the null hypothesis, and if $dw(u) = \alpha\exp(-\alpha^2u^2)du$ is chosen in the definition of $n\Delta_n$, then equations (7.2) may be used to compute cumulant matching approximations of the three types that have been described for the null asymptotic distribution of the test statistic. Results are presented in Table 1 for the case where $\gamma = \frac{\sigma}{\alpha} = 1$. Also included is the distribution of $n\Delta_n$ calculated by means of approximating quadratures as in example (c) of Section 5. Quadratures of order 48 and 64 yielded identical results for all entries in Table, which are thought to give the true distribution of $n\Delta_n$ to within $\pm$ 10^{-4}.

It can be seen that the distribution of the weighted sum of two chi-squared variates provides an excellent approximation for the asymptotic distribution of $n\Delta_n$, particularly for large x. The Pearson approximation gives results which are almost as good in the upper tail, although as must be expected it is not very accurate in the vicinity of the origin. The Pearson and sum of $2\chi^2$'s approximations appear to be adequate for most practical purposes, at least if the desired size of the test is not unusually large. Patnaik's approximation is inferior to Pearson's approximation, and is generally preferable since the latter requires only slightly more effort.

TABLE 1

Approximations to the Asymptotic Null Distribution of $n\Delta_n$ When the Underlying Population is Normal: $\gamma = \sigma/\alpha = 1$

x	Patnaik	Pearson	Sum of 2 χ^2's	$P(n\Delta_n \leq x)$
0.1	.1833	.1021	.1431	.1396
0.2	.3277	.3239	.3024	.3072
0.3	.4453	.4627	.4376	.4426
0.4	.5417	.5651	.5466	.5494
0.5	.6210	.6444	.6332	.6341
0.6	.6865	.7072	.7019	.7014
0.7	.7405	.7579	.7564	.7554
0.8	.7851	.7990	.7999	.7988
0.9	.8220	.8327	.8350	.8339
1.0	.8526	.8604	.8633	.8624
1.1	.8778	.8833	.8864	.8857
1.2	.8988	.9023	.9052	.9047
1.3	.9161	.9181	.9207	.9204
1.4	.9304	.9313	.9336	.9334
1.5	.9423	.9422	.9442	.9441
1.6	.9522	.9514	.9530	.9531
1.7	.9603	.9591	.9604	.9605
1.8	.9671	.9655	.9666	.9667
1.9	.9727	.9710	.9718	.9719
2.0	.9774	.9755	.9761	.9762
2.1	.9812	.9793	.9798	.9799
2.2	.9844	.9825	.9828	.9829
2.3	.9871	.9853	.9854	.9855
2.4	.9893	.9875	.9876	.9877
2.5	.9911	.9895	.9895	.9896
2.6	.9926	.9911	.9911	.9911
2.7	.9939	.9925	.9924	.9925
2.8	.9949	.9936	.9935	.9936
2.9	.9958	.9946	.9945	.9945
3.0	.9965	.9954	.9953	.9953

A similar comparison of approximations results when the underlying population is Cauchy with location parameter θ and scale parameter σ $dw(u) = \alpha \exp(-\alpha|u|)du$ and the ratio $\gamma = \sigma/\alpha$ was taken over a range of values including unity. The results were the same as for the normal care.

The cumulant matching approximations were also investigated when the distribution specified under the null hypothesis is the stable law of index ½ whose characteristic function is

$$\Phi_0(u) = \exp(-|u|^{\frac{1}{2}} (1-i \operatorname{sgn} u)).$$

The statistic considered was the discrete approximation to

$$n\Delta_n = n \int_{-\infty}^{\infty} |\hat{\Phi}_n(u) - \Phi_0(u)|^2 e^{-|u|}du$$

given by

$$n\Delta_n^{(p)} = n \sum_{k=1}^{p} |\hat{\Phi}_n(u_k) - \Phi_0(u_k)|^2 w_k$$

where u_k and w_k, $k = 1,2,\ldots, 60$ are the (two-sided) Laguerre integration abscissae and weights. The natural weight function is $\exp(-|u|^{\frac{1}{2}})$ but $\exp(-|u|)$ was used for computational convenience. The cumulants are obtained from (5.8). The agreement between the asymptotic null distribution given by Theorem 3.1 and the approximation is not quite as good as in the Gaussian and Cauchy examples but still the maximum deviation between the exact and approximate distribution determined by the weighted sum of two independent χ^2 variates is only .008 and in the upper decile is at most .003. Pearson's approximation again performs quite well in the upper tail and would probably be preferred due to its simplicity.

Our computations indicate that cumulant-matching approximations provide a suitable means for the evaluation of critical values required for use in tests of goodness-of-fit. Since such approximations do not require excessive computation or analysis, they make the use of the statistic $n\Delta_n$ feasible from a practical point of view over a wide range of situations.

8. RATE OF CONVERGENCE

Since the derivation of the finite sample distribution of the variate $n\Delta_n$ appears to be intractable, any inferences to be drawn through observation of this statistic will ordinarily be based on its asymptotic properties. Inferences could also be based on the finite sample cumulants (2.4) and (2.5), obtained in special cases by explicit integration or in general by numerical integration, and application of Patnaik's approximation. This would not be difficult in principle but could be avoided if convergence of the finite sample distributions is sufficiently rapid so that the statistic is of value for moderate sample sizes.

To see whether the null distribution of $n\Delta_n$ is well-approximated by its asymptotic distribution for moderate n, samples from each of the normal (γ=1), Cauchy (γ=1), and stable populations considered in the previous section were simulated. Empirical distribution functions of $n\Delta_n$ ($n\Delta_n^{(p)}$ for the stable case), based on either 2000 or 12000 replications, were generated for sample sizes n of 5, 15, 25, and 50.

Comparison of empirical and asymptotic distribution functions was made by the usual chi-squared statistic applied in each set of replications to test the hypothesis that it was drawn from the associated asymptotic distribution. The probability levels corresponding to these tests are given in Table 2. Based on this empirical evidence the finite sample size distribution of $n\Delta_n$ do seem to converge rapidly, and use of the asymptotic distribution when sample size is greater than or equal to 5, say, would appear to be justified.

9. Acknowledgement

The work of A.S. Paulson was supported in part by U.S. Army Research Office contract DAAG-29-81-K-0110.

TABLE 2

P-Value of the χ^2 Goodness-of-Fit Test of Asymptotic Distribution for Finite Sample Sizes n

Case	n = 5	15	25	50
Normal	.77	.72	.67	.46
Cauchy	.58	.94	.36	.30
Stable (½)	.32	.80	.61	.23

REFERENCES

Abramowitz, M. and Stegun, I. (1964). Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, National Bureau of Standards, Applied Matheamtical Series, 55.

Bryant, J.L. and Paulson, A.S. (1979). A Note on Characteristic Function-based Estimators. Sankhya, A, 41, pp. 109-116.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton: Princeton University Presss.

Csorgo, S. (1981). The Empirical Characteristic Process when Parameters are Estimated. In: Contributions to Probability: A Collection of Papers Dedicated to Eugene Lukacs, edited by J. Gani and V.K. Rohatgi, pp. 215-230.

Csorgo, S. (1984). Testing by the Empirical Characteristic Functions: A Survey. In: Proc. Third Progue Symp. Asymptotic Statistics, pp. 45-56, North-Holland, Auisterdoun, 1984.

Csorgo, S. (1986). Testing for Normality in Arbitrary Dimensionality. Annals of Statistics, 14, to appear.

Dunford, N. and Schwartz, J.T. (1958). Linear Operators, Part I: General Theory. New York: Interscience.

Durbin, J. and Knott, M. (1972). Components of Cramer-von Mises Statistics I, J.R. Statist. Soc. B, 34, pp. 290-307.

Feuerverger, A. and Mureika, R.A., (1977). The Empirical Characteristic Function and Its Applications, Annals of Statistics, 5, pp. 88-97.

Gil-Pelaez, J. (1951). Note on the Inversion Theorem, Biometrika, 38, pp. 481-482.

Gradshteyn, I.S. and Ryzhik, I.M. (1965). Tables of Integrals, Series, and Products. New York: Academic Press.

Heathcote, C.R. (1972). A Test of Goodness of Fit for Symmetric Random Variables, Austral. J. Statist., 14, pp. 172-181.

Hochstadt, H. (1972). Integral Equations. New York: Wiley.

Imhoff, P.J. (1961). Computing the Distribution of Quadratic Froms in Normal Variables, Biometrika, 48, pp. 419-426.

Koutrouvelis, I.A. (1980). A Goodness-of-Fit Test of Simple Hypotheses Based on the Empirical Characteristic Function. Biometrika, 67, pp. 238-240.

Koutrouvelis, I.A., and Kellermeier, J. (1981). A Goodness-of-fit Test Based on the Empirical Characteristic Function when Parameters must be Estimated. Journal of the Royal Statistical Society, B, pp. 173-1976.

Lukacs, E. (1970). Characteristic Functions. Darien: Hafner.

Lukacs, E. and Laha, R.G. (1964). Applications of Characteristic Functions. New York: Hafner.

Patnaik, P.B. (1959). The Non-Central χ^2 and F Distributions and Their Applications, Biometrika, 36, pp. 202-232.

Paulson, A.S. and Thornton, J. (1975). A New Approach to Goodness-of-Fit. In: Modeling and Simulation, Vol. 6, W. G. Vogt and M.H. Mickle, editors. Pittsburgh: The Instrument Society of America.

Pearson, E.S. (1959). Note on an Approximation to the Distribution of Non-Central χ^2, Biometrika, 46, p. 364.

Pogorzelski, W. (1966). Integral Equations and Their Applications, Vol. I. Warsaw: PWN - Polish Scientific Publishers.

Rao, C.R. (1965). Linear Statistical Inference and Its Applications. New York: John Wiley & Sons.

Schmeidler, W. (1965). Linear Operators in Hilbert Space. Translated from the German by J. Strum. New York: Academic Press.

Smithies, F. (1958). Integral Equations, Cambridge Tracts in Mathematics and Mathematical Physics, No. 49. Cambridge: Cambridge University Press.

Stroud, A. and Secrist, D. (1966). Gaussian Quadrature Formulas. New York: John Wiley and Sons.

MASTER COPY - FOR REPRODUCTION PURPOSES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AR[illegible]-16	2. GOVT ACCESSION NO.: N/A
3. RECIPIENT'S CATALOG NUMBER: N/A	
4. TITLE (and Subtitle): Goodness-of-Fit Based on Integrated Squared Discrepancies Between Sample and Population Characteristic Functions	5. TYPE OF REPORT & PERIOD COVERED: Working Paper
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): J. L. Bryant, A. S. Paulson	8. CONTRACT OR GRANT NUMBER(s): DAAG29-81-K-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Rensselaer Polytechnic Institute, Troy, New York 12180	10. PROGRAM ELEMENT PROJECT TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

sample characteristic function, goodness-of-fit, weighted sum of chi-squared variates, consistency, asymptotic distribution, rate of convergence, cumulants, approximate distribution.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We assess the feasibility of a goodness-of-fit test based on an integral of the weighted squared modulus of the discrepancy between the sample and population characteristic functions. The resulting statistic is therefore analogous to the Cramer-von Mises statistic and is shown to reduce to it as a special case. A number of approximations to the null distribution of the test statistic are considered and are found to be successful in simplifying its application without due loss of accuracy.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

4-87

DTIC